Phenotype-Genotype Integrator (PheGenI)

A tool that integrates the search and retrieval of associated genotype-phenotype data

https://www.ncbi.nlm.nih.gov/gap/PheGenI

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope and access

The Phenotype-Genotype Integrator (PheGenI) is a web interface that integrates data from various databases at NCBI with data from the National Human Genome Research Institute (NHGRI) Genome-wide Association Study (GWAS) Catalog [1, 9]. This phenotype-oriented resource can facilitate the prioritization of variants to follow up, study design consideration, and the generation of biological hypotheses. Searches can be conducted using chromosomal location, gene, SNP or phenotype.

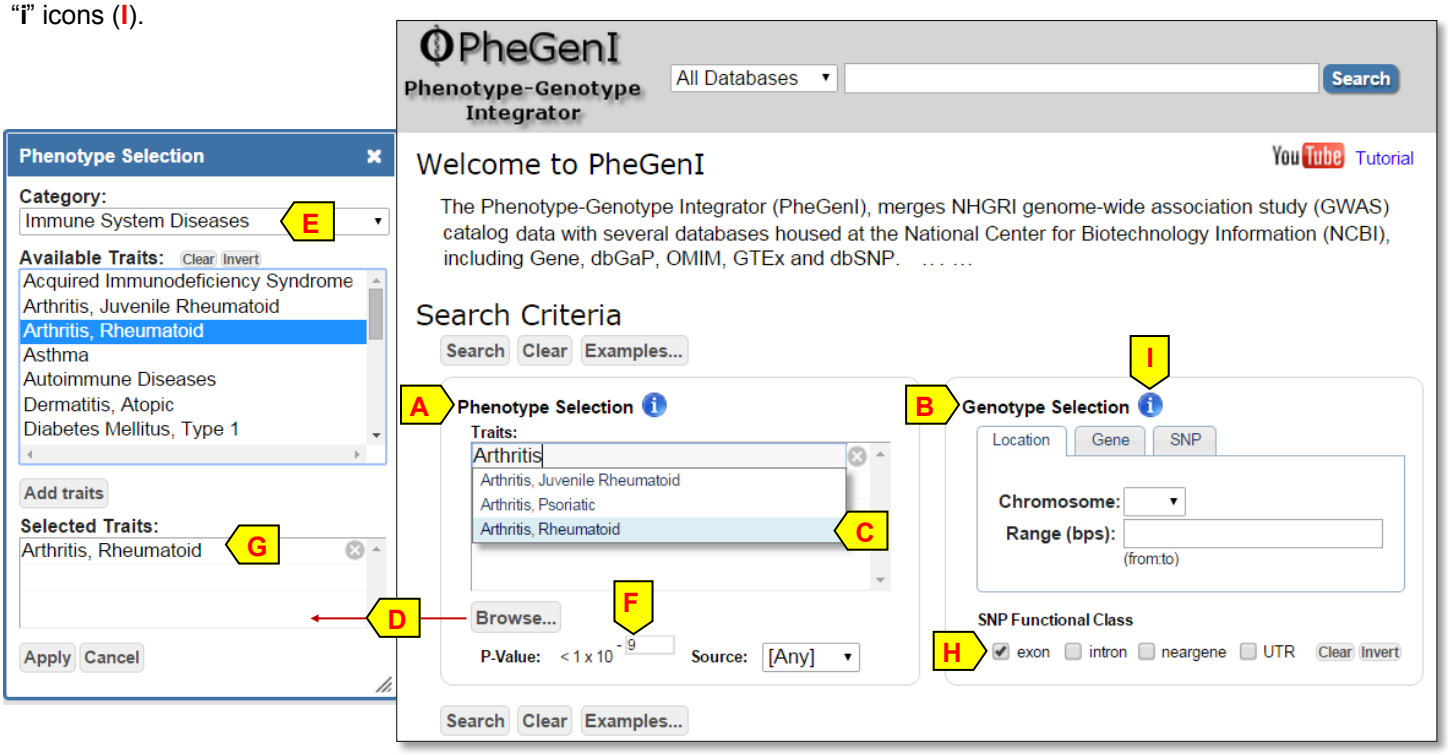
PheGenI supports downloads of search results and provides displays of results on the genome. Many of the genotype-phenotype associations have been extracted from publications by staff at NHGRI, with others from public data submitted to NCBI's dbGaP database [2]. In addition to GWAS results, PheGenI provides access to expression trait loci data from the GTEx [7] and links to Gene [3] as well as dbSNP [4].

PheGenI can be accessed directly using this URL:
www.ncbi.nlm.nih.gov/gap/PheGenI
A link to this tool is also available from the dbGaP homepage at:
www.ncbi.nlm.nih.gov/gap/

Searching using criteria selection

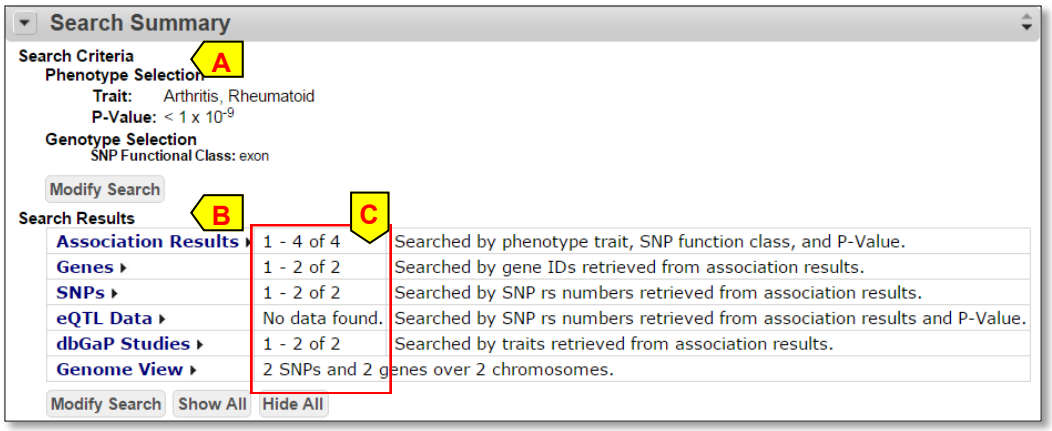
PheGenI provides two types of searches, Phenotype Selection (**A**) and Genotype Selection (**B**). The Phenotype Selection searches for genotypes associated with certain phenotypes (diseases or other traits) of interest. The Genotype Selection searches with a genotypic query (genomic location, gene or SNP ID) to find associated phenotypes. Selections from both types can be combined to get more specific results.

Phenotypic traits can be entered into the "Phenotype Selection" box followed by selection from the suggested list (**C**), or they can be selected from the Phenotype Section popup activated by the Browse button (**D**). Selecting an entry from the Category field (**E**) further narrows down the trait lookup. A P-Value (**F**) filter is provided for selection of genotype data from GWAS studies with statistical significance higher than the specified threshold. The Genotype Selection allows the examination of reported phenotypes for a selected genomic regions, genes of interest, or SNPs (represented by rsIDs, such as rs328) of interest. The example search uses Arthritis, Rheumatoid (**G**) as the trait from a broader category of the Immune System Diseases in the Phenotype Selection. The search is further restricted to exonic SNPs (**H**) using the Genotype Selection. This combination of criteria retrieves arthritis-associated exonic variants reported by GWAS. Existing input selections can be reset using the Clear button and context-specific help information is available by clicking the "**i**" icons (**I**).



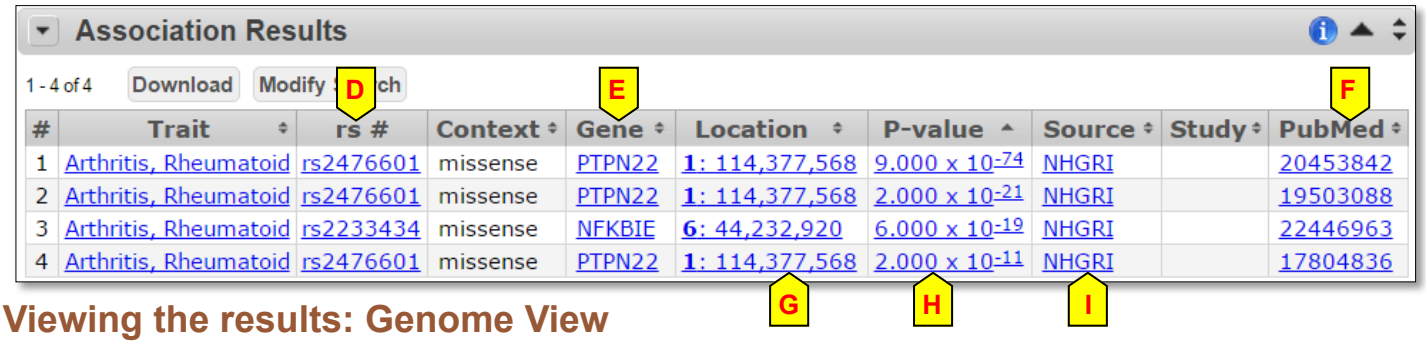
Viewing the results: Summary

The search results page has several collapsible sections. The display below is the Search Summary section, which highlights the search criteria (A) and results (B) with hyperlinks to corresponding sections underneath. This example (C) indicates that the search retrieves four association studies identifying two different genes and a total of two SNPs for the phenotype and type of variations specified.



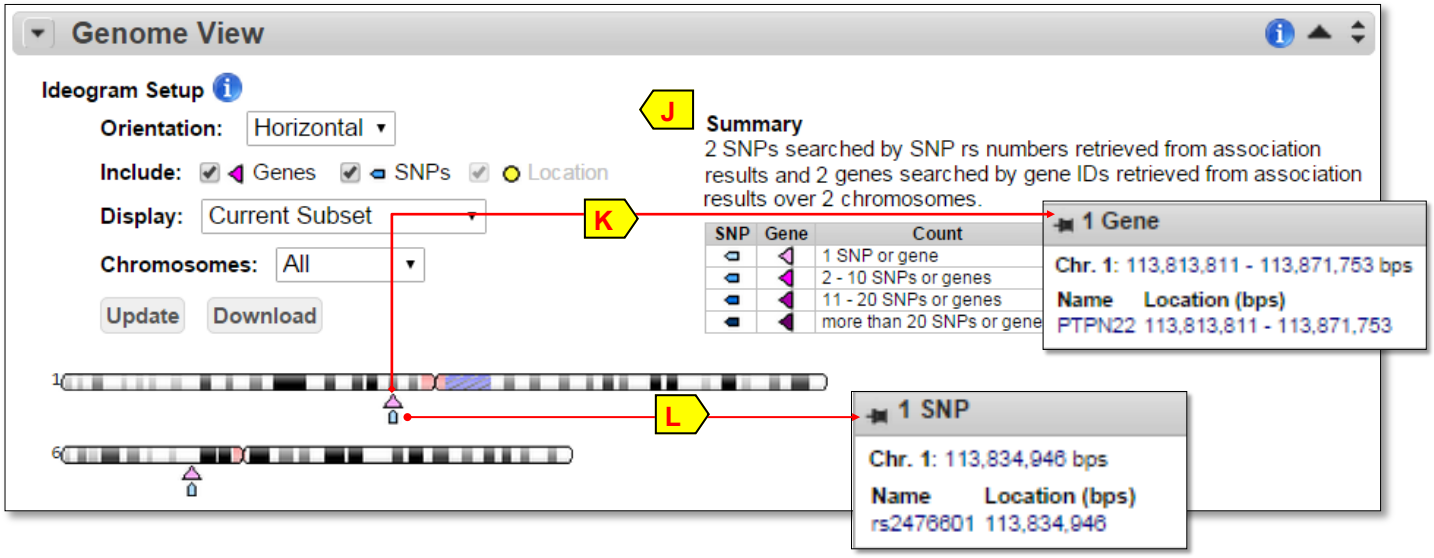
Viewing the results: Association Results

The association studies retrieved by the search are displayed in the Association Results section (partially shown below). The example indicates that the same SNP (rs2476601, D) from the PTPN22 gene (E) is associated with rheumatoid arthritis in three different GWAS studies, as indicated by three different PubMed IDs (F). In these studies, this exonic SNP, specified by input search criteria, showed a strong association with the disease as indicated by their low P-values (G). One other SNP associated with a different gene (NFKBIE) was also found to be associated with this trait. This table provides links to relevant records in other NCBI databases. The rsIDs in the rs # column (D) link to the full reports of corresponding SNPs, and the symbols in the Gene column (E) link to the NCBI Gene records. Coordinates in the Location column (G) link to graphical displays of the genomic region in the Sequence Viewer [5]. The P-value (H) links to a display in the GaP Browser [6] for interactive exploration of the genotype-phenotype association. For this sample set of results, the Source (I) of the association information is the NHGRI GWAS Catalog.



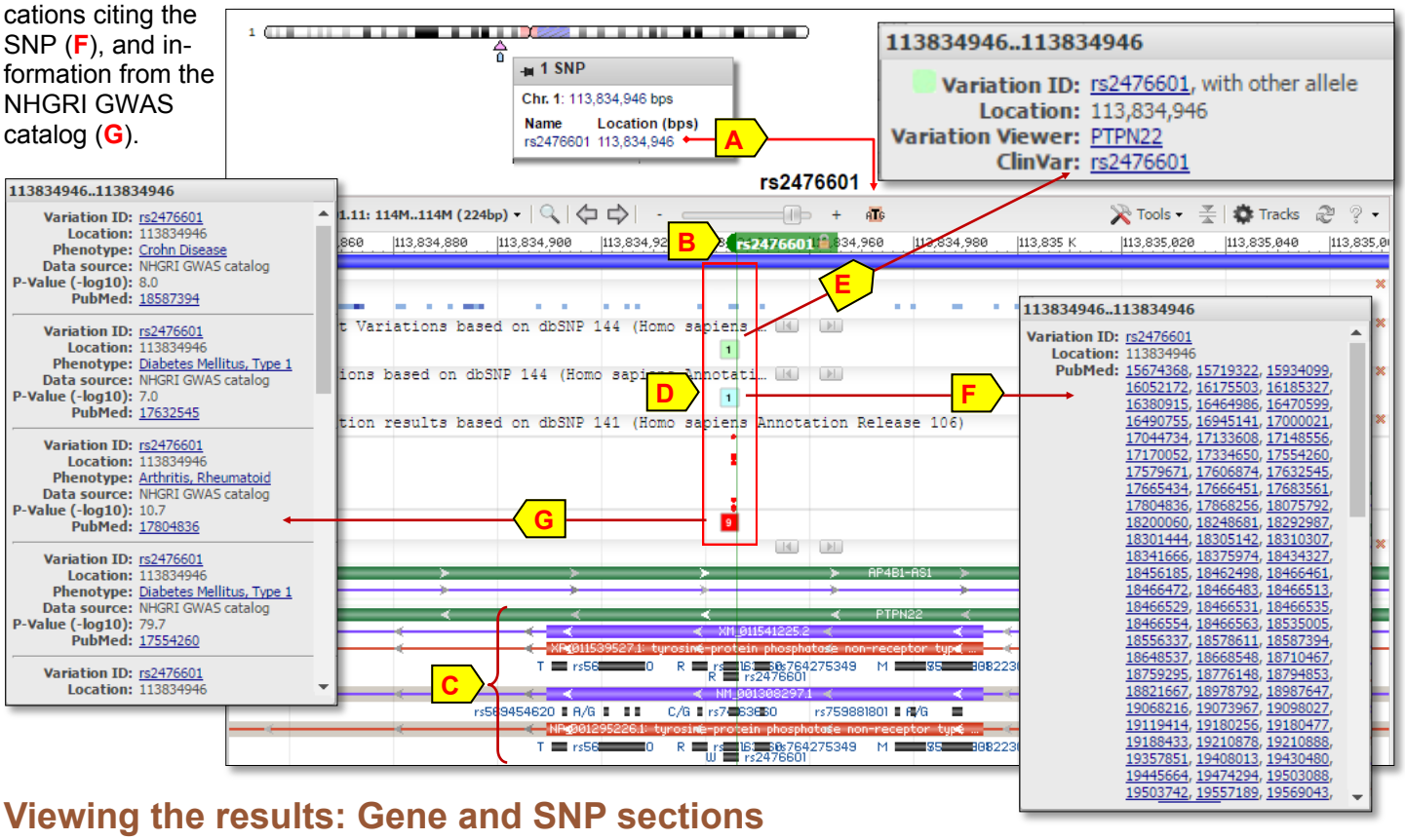
Viewing the results: Genome View

The Genome View section (below) has customizable display options and enables a summary view of the Genes and SNPs positions on their chromosomes. The color-coded arrows (J) indicate the locations of genes and SNPs. A short summary is displayed when the mouse pointer is placed over one of the markers (K, L). Clicking on a marker adds or updates the graphical display in the Genome View section below to provide additional details.



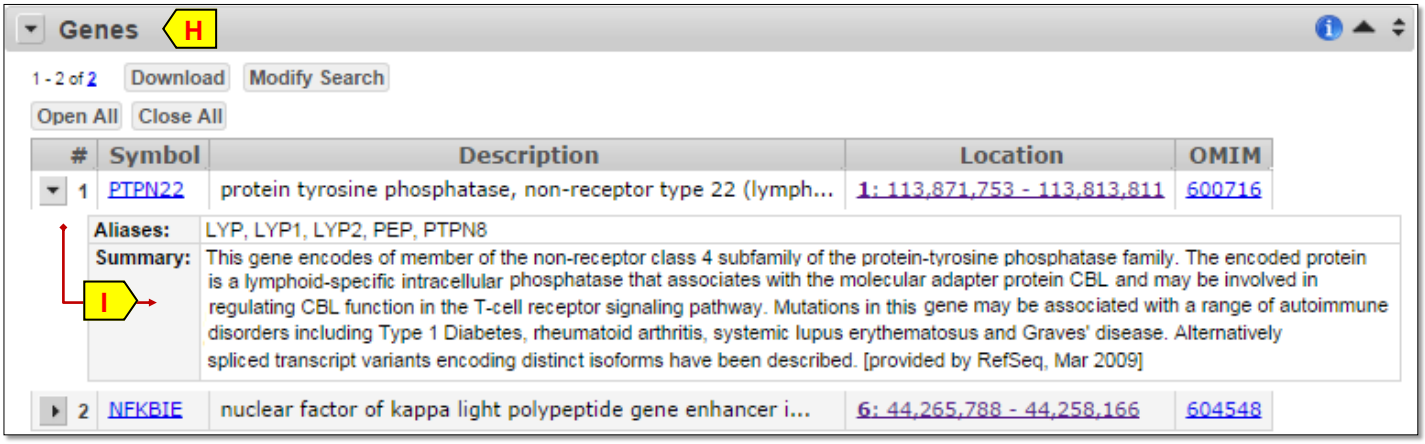
Viewing the results: Genome View (cont.)

The graphical view enables detailed examination of a specific SNP (rs2476601) activated by clicking the coordinate in the popup (**A**). In this display, the position of this SNP is indicated by a marker (**B**) with the exon bearing the SNP given below (**C**). Hovering over boxed numbers (**D**) in different variation tracks provide information from the ClinVar (**E**), publications citing the SNP (**F**), and information from the NHGRI GWAS catalog (**G**).

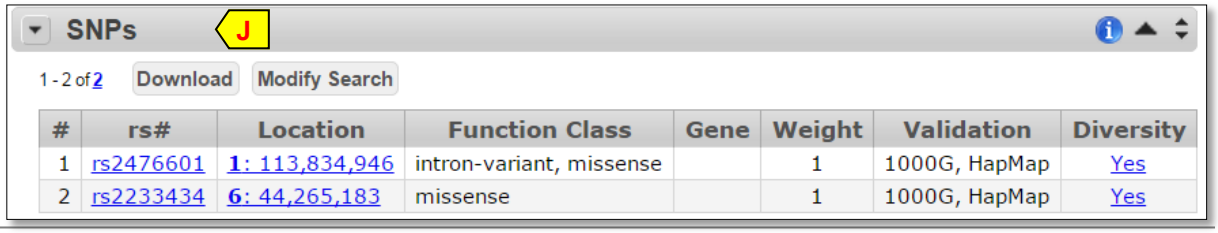


Viewing the results: Gene and SNP sections

The Genes section of the PheGenI report provides the list of phenotype-associated genes in a table (**H**), showing their official symbol, names, the genomic coordinates, and relevant OMIM [8] entries. Clicking an arrow (**I**) to the left reveals the summary of that gene in a section expanded below the row.



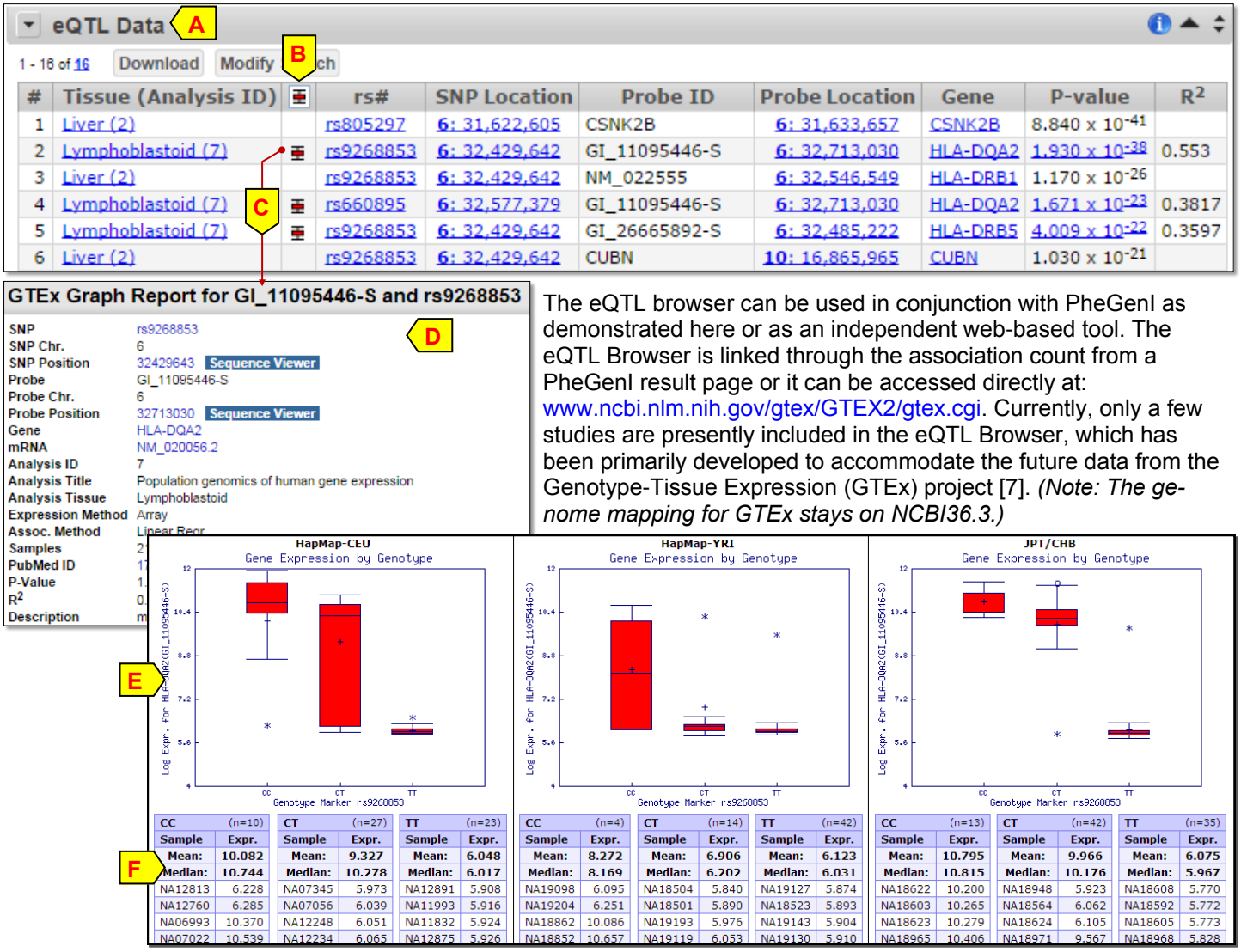
The SNPs section (**J**) tabulates the key characteristics of phenotype-associated SNPs. It lists the genomic location, the functional class, and associated gene for these SNPs. Entries in the Genes and SNPs tables are hyperlinked to records in other databases or to specialized graphical displays in Sequence Viewer to provide additional details.



#	rs#	Location	Function Class	Gene	Weight	Validation	Diversity
1	rs2476601	1: 113,834,946	intron-variant, missense		1	1000G, HapMap	Yes
2	rs2233434	6: 44,265,183	missense		1	1000G, HapMap	Yes

Viewing the results: eQTL Data

The term eQTL stands for "expression quantitative trait loci." In the example search described earlier, the "eQTL Data" section has no data. This is not surprising as variants that affect gene expression are most often located in the regulatory regions of the genes and thus not in exonic regions. Modifying the original search by removing the exon restriction (using the "Modify Search" button) produces a much larger number of associations with a data-containing "eQTL Data" section (**A**). Icons (**B**) in the third column of the table indicates expression data is available from eQTL. Clicking on an icon (**C**) displays the report on gene expression by genotype from the eQTL browser. The report contains a summary at the top (**D**). The graphs (**E**) sum up the differences in level of expression of the target gene for different genotypes from the same populations, with detailed expression data for individuals given in tables below (**F**).



#	Tissue (Analysis ID)	rs#	SNP Location	Probe ID	Probe Location	Gene	P-value	R^2
1	Liver (2)	rs805297	6: 31,622,605	CSNK2B	6: 31,633,657	CSNK2B	8.840×10^{-41}	
2	Lymphoblastoid (7)	rs9268853	6: 32,429,642	GI_11095446-S	6: 32,713,030	HLA-DQA2	1.930×10^{-38}	0.553
3	Liver (2)	rs9268853	6: 32,429,642	NM_022555	6: 32,546,549	HLA-DRB1	1.170×10^{-26}	
4	Lymphoblastoid (7)	rs660895	6: 32,577,379	GI_11095446-S	6: 32,713,030	HLA-DQA2	1.671×10^{-23}	0.3817
5	Lymphoblastoid (7)	rs9268853	6: 32,429,642	GI_26665892-S	6: 32,485,222	HLA-DRB5	4.009×10^{-22}	0.3597
6	Liver (2)	rs9268853	6: 32,429,642	CUBN	10: 16,865,965	CUBN	1.030×10^{-21}	

SNP	rs9268853
SNP Chr.	6
SNP Position	32429643 Sequence Viewer
Probe	GI_11095446-S
Probe Chr.	6
Probe Position	32713030 Sequence Viewer
Gene	HLA-DQA2
mRNA	NM_020056.2
Analysis ID	7
Analysis Title	Population genomics of human gene expression
Analysis Tissue	Lymphoblastoid
Expression Method	Array
Assoc. Method	Linear Regr
Samples	2
PubMed ID	17
P-Value	1.
R^2	0.
Description	m

The eQTL browser can be used in conjunction with PheGenI as demonstrated here or as an independent web-based tool. The eQTL Browser is linked through the association count from a PheGenI result page or it can be accessed directly at: www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi. Currently, only a few studies are presently included in the eQTL Browser, which has been primarily developed to accommodate the future data from the Genotype-Tissue Expression (GTEx) project [7]. *(Note: The genome mapping for GTEx stays on NCBI36.3.)*

CC (n=10)		CT (n=27)		TT (n=23)	
Sample	Expr.	Sample	Expr.	Sample	Expr.
Mean:	10.082	Mean:	9.327	Mean:	6.048
Median:	10.744	Median:	10.278	Median:	6.017
NA12813	6.228	NA07345	5.973	NA12891	5.908
NA12760	6.285	NA07056	6.039	NA11993	5.916
NA06993	10.370	NA12248	6.051	NA11832	5.924
NA07022	10.539	NA12234	6.065	NA12875	5.926

CC (n=4)		CT (n=14)		TT (n=42)	
Sample	Expr.	Sample	Expr.	Sample	Expr.
Mean:	8.272	Mean:	6.906	Mean:	6.123
Median:	8.169	Median:	6.202	Median:	6.031
NA19098	6.095	NA18504	5.840	NA19127	5.874
NA19204	6.251	NA18501	5.890	NA18523	5.893
NA18862	10.086	NA19193	5.976	NA19143	5.904
NA18852	10.657	NA19119	6.053	NA19130	5.910

CC (n=13)		CT (n=42)		TT (n=35)	
Sample	Expr.	Sample	Expr.	Sample	Expr.
Mean:	10.795	Mean:	9.966	Mean:	6.075
Median:	10.815	Median:	10.176	Median:	5.967
NA18622	10.200	NA18948	5.923	NA18608	5.770
NA18603	10.265	NA18564	6.062	NA18592	5.772
NA18623	10.279	NA18624	6.105	NA18605	5.773
NA18965	10.406	NA18971	9.567	NA18968	5.828

References and documents

1. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Hindorff LA, et. al. (2009). Proc Natl Acad Sci USA.106(23): 9362-7.
2. The NCBI dbGaP database of genotypes and phenotypes. Mailman MD, et. al. (2007). Nat Genet. 39(10):1181-6.
3. Entrez Gene: gene-centered information at NCBI. Maglott D, et al (2011). Nucleic Acids Res. 39(Database issue):D52-7.
4. dbSNP: the NCBI database of genetic variation. Sherry ST, et. al (2001). Nucleic Acids Res. 2001 Jan 1;29(1):308-11.
5. Sequence Viewer homepage: www.ncbi.nlm.nih.gov/projects/sviewer/
6. GaP Browser link: www.ncbi.nlm.nih.gov/projects/SNP/GaPBrowser_prod/callGaPBrowser2.cgi
7. GTEx link: www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi
8. OMIM site: www.omim.org/
9. Additional information on GWAS: http://www.genome.gov/26525384
10. Additional facts on PheGenI: www.genome.gov/27543987
11. PheGenI video tutorial: www.youtube.com/ncbinlm#p/c/8FD4CC12DABD6B39/1/XS3p924nWCA